Intron-regulated expression of $SUVH3$, an Arabidopsis $Su(var)3$-$9$ homologue

Juan A. Casas-Mollano, Nga T. Lao and Tony A. Kavanagh*

Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland

Received 27 April 2006; Accepted 22 June 2006

Abstract

SU(VAR)3-9 proteins are key regulators of heterochromatin structure and function in plants, mammals, Drosophila, and yeast. In contrast to animals and fungi, plants contain numerous $Su(var)3$-$9$ homologues ($SUVH$), the members of which form a discrete subfamily. The SU(VAR)3-9 and SUVH proteins associate with heterochromatin and possess histone methyltransferase activity, indicating that they participate in the organization of transcriptionally repressive chromatin. The Arabidopsis thaliana genome contains 10 SUVH genes, belonging to four phylogenetically distinct subgroups: SUVH1, SUVH2, SUVH4, and SUVH5. The structure and expression of SUVH3, a member of the SUVH1 subgroup was investigated. SUVH3 was shown to be broadly expressed during plant development with the highest levels found in proliferating cells. The encoded protein localized in subnuclear foci and remained associated with condensed chromosomes throughout mitosis. A deletion analysis of the SUVH3 upstream region further revealed that an intron located in the 5' UTR is a key regulator of strong and constitutive $SUVH3$ expression.

Key words: Arabidopsis, gene expression, intron, $Su(var)3$-$9$, SUVH3, 5' UTR.

Introduction

The establishment and maintenance of transcriptionally permissive and repressive chromatin domains, euchromatin and heterochromatin, respectively, is essential for the correct temporal and spatial regulation of gene expression during development. These epigenetically stable forms of chromatin are established by differential covalent modification of the highly conserved N-terminal tails of the core histones via methylation, acetylation, phosphorylation, ubiquitination, or ADP ribosylation of specific amino acid residues (Jenuwein and Allis, 2001). The modified histones in turn recruit chromatin-specific regulatory complexes that determine the transcriptional potential of eu- and heterochromatic regions throughout the genome as originally proposed in the ‘histone code’ hypothesis (Strahl and Allis, 2000). Differential methylation of histone lysine (K) residues by SET domain proteins has been shown to play a key role in regulating the dynamic balance between alternative chromatin states (Ebert et al., 2004). Thus, euchromatin is characterized by histone H3 methylated at K4, K36, and K79 while heterochromatin contains methylated H3K9, H3K27, and histone H4 methylated at K20.

The SET domain is a conserved ~130 amino acid sequence motif originally identified at the C-terminal end of proteins encoded by the Drosophila genes $Su(var)3$-$9$, $E(z)$, and Trithorax (Tschiersch et al., 1994). Proteins containing a SET domain can be grouped into four subfamilies based on the extent of sequence conservation (Jenuwein et al., 1998). The SET domain of several members of the $Su(var)3$-$9$ subfamily have been shown to possess H3K9 methyltransferase activity (Rea et al., 2000) and to function, in collaboration with DNA methyltransferases, in the organization of repressive heterochromatin (Stancheva, 2005). The genome of Arabidopsis thaliana contains 10 genes that encode SET domain proteins classed as $Su(var)3$-$9$ homologues ($SUVH$) in marked contrast to animals and fungi which have one or two, respectively (Baumbusch et al., 2001). Multiple SUVH homologues have also been identified in tobacco, rice, and maize (Baumbusch et al., 2001; Shen, 2001; Springer et al., 2003; the Plant Chromatin database). The encoded proteins (Class V; Springer et al., 2003) possess a domain architecture that is similar, but not identical to Drosophila SU(VAR)3-9, characterized by the presence of a SET domain flanked by cysteine-rich domains termed the pre-SET and post-SET.

* To whom correspondence should be addressed. E-mail: tkvanagh@mail.tcd.ie

© The Author [2006]. Published by Oxford University Press [on behalf of the Society for Experimental Biology]. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org
Uniquely, plant SUVH proteins contain an additional domain designated SRA (SET and RING finger associated) (Baumbusch et al., 2001) that is thought to play a role in targeting SUVH proteins to specific chromatin subdomains (Citterio et al., 2004; Yu et al., 2004).

Investigations of null mutants and ectopic overexpression lines indicate that SUVH2 (Naumann et al., 2005; Fischer et al., 2006) and SUVH4 (also known as KRYPTONITE) (Jackson et al., 2002; Malagnac et al., 2002) play major roles in directing epigenetic modifications in Arabidopsis. In each case, the encoded protein has been shown to possess methyltransferase activity targeting H3K9 (and H4K20 in the case of SUVH2) and to accumulate in nuclei where it associates with repressive heterochromatin (Yu et al., 2004; Naumann et al., 2005; Fischer et al., 2006). Loss of SUVH2 substantially alters all heterochromatin-specific histone methylation marks (Naumann et al., 2004) while loss of SUVH4 activity causes a large reduction in the levels of H3K9 di-methylation (Jackson et al., 2004). By contrast, the loss of SUVH1 (Naumann et al., 2005), SUVH5 (Ebbs and Bender, 2006) or SUVH6 (Ebbs et al., 2005) results in only minor reductions in global H3K9 methylation levels.

SUVH genes appear to be expressed in most organs in Arabidopsis, tobacco, maize, and rice (Baumbusch et al., 2001; Shen, 2001; Springer et al., 2003; The Plant Chromatin database), as determined by RT-PCR and northern blot analyses. These analyses, however, do not reveal whether the expression of individual SUVH genes is truly constitutive or varies in a tissue- or cell-specific manner during growth and development. A comparative analysis of the genomic and cDNA sequences of SUVH3 revealed the presence of a 5'-UTR containing two introns followed by a single large uninterrupted open reading frame (ORF). Increasing evidence indicates that introns, particularly those located in the 5'-ORF, can profoundly affect gene expression patterns.

Introns can regulate gene expression by enhancing transcription initiation, by increasing mRNA accumulation co- or post-transcriptionally, or by enhancing translational efficiency (LeHir et al., 2003). The second intron of the Arabidopsis AGAMOUS (AG) gene, for example, contains cis-acting DNA sequences that bind the transcription factors LEAFY and WUSCHEL which control floral fate and floral-specific expression of AG, respectively (Busch et al., 1999; Lohmann et al., 2001). Introns that act co- or post-transcriptionally to increase gene expression, a process referred to as intron-mediated enhancement (IME), have been documented in several plant genes, including the maize genes Alcohol dehydrogenase1 and Bronze1 (Callis et al., 1987), Shrunken1 (Clancy and Hannah, 2002), and the Arabidopsis gene Phosphoribosylanthranilate transferase1 (Rose and Last, 1997).

Here, it is shown that the protein encoded by Arabidopsis SUVH3 accumulates in discrete foci in interphase nuclei and associates with chromosomes throughout the cell division cycle. It is also shown that SUVH3 is broadly expressed during plant development and that strong and constitutive expression of SUVH3 in Arabidopsis is dependent on the first of two introns located in the 5'-UTR.

**Materials and methods**

**Plant material**

Arabidopsis thaliana ecotype Columbia with the glabra-1 marker (Catalogue number: WT-1A) was obtained from LEHLE SEEDS (Texas, USA). Nicotiana tabacum cv. Bright Yellow 2 (BY2) cell suspensions were maintained as described by Banjoko and Trelease (1995).

**Nucleotide and amino acid sequence analysis**

Nucleotide sequence alignments and oligonucleotide primer design was carried out using Vector NTI software (InforMax Inc.). Searches for cis-regulatory elements in promoter sequences were performed in the PLACE database (Plant Cis-acting Regulatory DNA Elements: http://www.dna.affrc.go.jp/htdocs/PLACE). Protein domains were identified using the SMART database (Simple Modular Architecture Research Tool: http://smart.embl-heidelberg.de), the MOTIF program (http://www.motif.genome.ad.jp), and the PROSITE database (http://www.expasy.ch/prosite). PSORT prediction (http://psort.nibb.ac.jp/form.html) was used to predict protein subcellular location and nuclear localization signals. Putative serine, threonine, and tyrosine phosphorylation sites were identified using the NetPhos 2.0 Prediction Server (http://www.cbs.dtu.dk/services/NetPhos).

**Molecular cloning and sequencing**

DNA cloning process was carried out according to standard methods (Sambrook et al., 1989). cDNA sequencing and all sequencing to verify the integrity of gene constructs were performed by MWG Biotech AG (Germany).

**Identification of a full-length SUVH3 cDNA**

The genomic sequence of SUVH3 gene was initially identified on the basis of the similarity between the encoded ORF and the C-terminal end of the SET domain of the Drosophila Su(var)3-9 gene. It was then used to search in the Arabidopsis EST database (http://www.arabidopsis.org/Blast) and two available cDNAs were obtained from the Arabidopsis Biological Resource Center (ABRC). Plasmid DNA was isolated and the ends of the cDNA inserts sequenced using M13 universal primers. From two requested cDNAs, only T04123 was shown to contain a complete SUVH3 ORF insert and was then fully sequenced.

**Construction of a GFP-SUVH3 translational gene fusion**

To create an N-terminal translational fusion between GFP and SUVH3 a GFP (719 bp) fragment was amplified from the plasmid psmGFP (Davis and Vierstra, 1998) using the primers GFPXbaF (5'-ctctgtagCAACAATGATGGGAGAGAAC; the lower case bases were introduced in order to create a Xbal site) and GFPKpnR (5'-ccgagctcTCATCCGAATGACCGC; the lower case bases were introduced in order to create a KpnI site). The SUVH3 ORF (2010 bp) was amplified from the full length cDNA (T04123) using Pfu DNA polymerase and the primers MACorfF (5'-gggggtaccATGCACCAATGATGGGAGAGAAG; the lower case bases were introduced in order to create a Xbal site) and MACorfR (5'-ccgagctcTCATCCGAATGACCGC; the lower case bases were introduced in order to create a SacI site). The resulting PCR fragments were cleaved with the appropriate restriction enzymes and were simultaneously

---

**Downloaded from** https://academic.oup.com/jxb/article-abstract/57/12/3301/452512 by guest on 14 February 2019
cloned into the XhoI and SacI sites of pUC19 to generate a GFP-SUVH3 translational fusion. The GFP-SUVH3 fragment was then subcloned between the double CaMV 35S promoter and the NOS terminator in the binary vector pROK-2CaMV, resulting in pGFP-SUVH3.

**Promoter-GUS fusions and intron deletion derivatives**

A 2032 bp DNA fragment comprising (5′-3′) 800 bp from the 3′ end of the upstream gene (At1g73090), the SUVH3 promoter region, the 5′-UTR and the first five codons of the SUVH3 ORF was amplified using the primers MACF (5′-ggctcagaCCCGTAGTTCAACCC-TTG; the lower case bases were added in order to create a XhoI site) and MACR (5′-ggggatctCCTCGTCTATACAGTAG; the lower case bases were added to create a BamHI site). The fragment was then cloned into the XhoI and BamHI sites of pGUS1, a pUC19 derivative containing the β-glucuronidase (GUS) reporter gene (Jefferson et al., 1987; TA Kavanagh, unpublished data). The introduced BamHI site was designed to facilitate an in-frame fusion between SUVH3 and the start codon of the GUS gene. The resulting plasmid pMAC-GUS1 was used to generate a set of deletion derivatives, as follows:

**DEL1** is a MAC-GUS derivative in which intron 1 has been deleted. The deletion was generated by inverse-PCR using pMAC-GUS1 as the DNA template and the primers Del1F (5′-ggctcaga-GAATTATCACACAAGGTCACTG; lower case bases were added in order to create a XhoI site) and Del1R (5′-gggtcagagGATGATGAAAGACAGAGAC; lower case bases were added in order to create a XhoI site). These primers were designed to amplify the DNA regions flanking intron 1 and the vector itself. The resulting PCR product was digested with XhoI and self-ligated to generate pGUS1-DEL1 which contains the SUVH3 promoter without intron 1 and two nucleotide changes in the junction between exon 1 and 2 caused by the engineered XhoI site.

**DEL2** was constructed by PCR-mediated deletion of the second intron using pMAC-GUS1 as the template and the primers MACF (sequence shown above) and Del2R (5′-ggggatctAGAAGCCCTTCCTCTATATAAGGAAG; lower case bases were added in order to create a BamHI site and the asterisk marks the junction between exon 2 and 3). The primer Del2R was designed to skip the sequence of intron 2 and leaving only the junction between exons 2 and 3. The resulting PCR product was cloned into the XhoI and BamHI sites of the pGUS1 vector. The BamHI site was positioned in the same place as in MACR to create a translational fusion between SUVH3 ATG and GUS.

In DEL1-2, sequences extending from +44 bp relative to the transcriptional start site (5′) to the fifth codon of SUVH3 were deleted by PCR amplification of the pMAC-GUS1 template using the oligonucleotide primers MACF and DEL1-2R (5′-ggggatctAGAAGCCCTTCCTCTATATAAGGAAG; lower case bases were added in order to create a BamHI site). The resulting PCR product was cloned into the XhoI and BamHI sites of the pGUS1 vector. The BamHI site was positioned in the same place as in MACR to create a translational fusion between SUVH3 ATG and GUS.

**DEL4** was constructed by deleting the entire SUVH3 promoter region extending upstream from +6 relative to the transcription start site. The SUVH3 promoter sequences were replaced by the −60 to +100 CaMV 35S minimal promoter (Campisi et al., 1999) generated by using the MACR reverse primer and two overlapping forward primers: 60Del4F1 (5′-cctctagaCCACCATATCTCCGCAAGAGCCCTTCCTCTATATAAGGAAG) and 60Del4R2 (5′-cctctcctCAATTAAAGGAAAGTCCTATCTGCCGATGCAGAG) (lower case bases were added in order to create a BamHI site). The underlined bases correspond to the sequence of the 5′-promoter and the italicized bases indicate the overlapping region between both primers. The resulting PCR product was cloned into the XhoI and BamHI site of the pGUS1 vector.

**DEL5** carries a deletion of the adjacent upstream gene sequences (At1g73090) included in pMAC-GUS1. pDEL5 was obtained by subcloning a Sph1-BamHI fragment of pMAC-GUS1 into the XhoI and BamHI site of the pTAk1 vector (Jefferson et al., 1987). The full-length MAC-GUS fusion and the various DEL derivatives were subcloned from pGUS1 into the XhoI and SacI sites of the binary vector pTAK1 (Jefferson et al., 1987). The resulting vectors were called pMAC-GUS, pDEL1, pDEL2, pDEL1-2, pDEL4, pDEL5 according to the deletion they carried.

**Transformation of Arabidopsis**

The various pTAK binary plasmid derivatives were transferred by electroporation into Agrobacterium tumefaciens strain AGL1 and the resulting Agrobacterium strains used to transform Arabidopsis plants via the ‘floral dip’ method (Clough and Bent, 1998).

**Stable transformation of tobacco BY2 cells**

A 4 ml aliquot of a 3-d-old BY2 cell culture, supplemented with 20 μM acetoxyringsone (added before transformation), was cocultivated with 200 μl of an overnight Agrobacterium culture in a 10 cm Petri dish plates for 2 d at 28 °C in darkness. After incubation cells were washed in 50 ml of BY2 medium supplemented with 500 mg l⁻¹ carbenicillin and then centrifuged at 1000 rpm for 4 min at room temperature in a swinging bucket rotor to pellet the cells. The cell pellet was spread in semi-solid BY2 medium supplemented with 100 mg l⁻¹ kanamycin and 500 mg l⁻¹ carbenicillin.

Transformed cells appeared after 3–4 weeks as small calli on plates and were maintained as a callus by transferring to a fresh plate once a month. From transformed callus a suspension culture was initiated by addition of a small callus clump to 20 ml of culture medium. Once a suspension was established, this was diluted in 50 ml. Eventually, transgenic cells in suspension grew as fast as wild-type BY2 cells and were subcultured once a week.

**Fluorometric GUS assay**

Seedlings were homogenized in 200 μl of extraction buffer (50 mM NaH2PO4, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1% sodium laurylsarcosine, 10 mM β-mercaptoethanol). Protein concentrations were determined by the dye-binding method of Bradford (Bradford, 1976). Fluorogenic reactions were carried out using 10 μl of the leaf extract in 100 μl of extraction buffer containing 0.5 mM MUG (4-methylumbelliferyl-β-D-glucuronide). Reactions were incubated at 37 °C for 3 h. Reactions were stopped by adding 900 μl of 0.2 M Na2CO3. Fluorescence was measured using a luminescence spectrometer Perkin Elmer LS-50B with excitation at 365 nm, emission at 455 nm, and slit widths set at 5 nm.

**Histochemical detection of GUS activity**

Small organs (flower and siliques), tissue pieces (leaves, roots and stems), or whole seedlings were immersed in GUS staining buffer containing 50 mM NaH2PO4, pH 6.8, 1 mM EDTA, 0.1% Triton X-100, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide, and 1 mM 5-bromo-4-chloro-3-indolyl-glucuronide (X-Gluc) (Jefferson et al., 1987), followed by vacuum-infiltration for 45 min. The reactions were incubated for 3–16 h at room temperature or 37 °C. Stained samples were bleached by incubation in 75% ethanol for several hours. Samples were observed in a Olympus SZX9 stereo microscope and a Olympus BX40 compound microscope. Pictures were taken with an Olympus DP10 digital camera.

**Fluorescence microscopy**

For fluorescence microscopy, Arabidopsis tissue pieces (leaves and roots) or tobacco BY2 cells were transferred to a glass slide and analysed in a Nikon Eclipse E600 microscope with a Y-LF Epifluorescence attachment. Pictures were taken with a Spot RT CCD colour digital camera.
Results

SUVH3 gene structure

SUVH3 (At1g73100), originally identified in a BLASTP survey of the Arabidopsis SET domain gene complement (Baumbusch et al., 2001), comprises an uninterrupted ORF (2007 bp) flanked by a 138 bp 5′ untranslated region (5′-UTR) and a 145 bp 3′-UTR. The putative ATG start codon is flanked by sequences that conform to the Kozac consensus for initiation of translation (Kozac, 1999). Comparison of a full-length cDNA from RIKEN and two EST sequences with the SUVH3 genomic sequence revealed the presence of two introns in the 5′-UTR, as included in the current TAIR gene model. Intron 1 (464 bp) is located at –220 bp relative to the start codon while intron 2 (111 bp) is located at –10 bp (Fig. 1). The AU content of the first and second intron is 65.3% and 67.6%, respectively.

The SUVH3 protein shows subnuclear localization and associates with chromosomes during cell division

SUVH3 encodes a putative protein of 669 amino acids with a predicted molecular weight of 73.4 kDa. The encoded protein contains an AT-hook domain, an SRA domain, and a C-terminal SET domain, with flanking pre-SET and post-SET domains (Fig. 2). Further analysis revealed the presence of a putative bipartite nuclear localization signal (NLS) immediately downstream of the AT-hook domain and several potential serine and threonine phosphorylation signals flanking the NLS (Fig. 2).

The subcellular localization of the SUVH3 protein was investigated by constructing a translational gene fusion between SUVH3 and GFP and using it to produce stably transformed Arabidopsis plants and N. tabacum BY2 cell lines. In transgenic plants (data not shown) and BY2 cell expressing GFP-SUVH3, green fluorescence was localized exclusively in the nuclei (Fig. 2). The green fluorescence was not uniformly distributed within nuclei, but rather localized in speckles some of which were associated with nucleoli. During mitosis, the GFP-SUVH3 fusion protein was strongly associated with chromosomes throughout the mitotic phases: prophase, metaphase, anaphase, and telophase (Fig. 2). Although several putative phosphorylation sites flank the NLS, regulated nuclear localization, as shown by nucleo-cytoplasmic partitioning of GFP-SUVH3, was not observed in any organ or tissue analysed. This indicates that SUVH3 is a nuclear protein and its NLS functions as a constitutive signal for nuclear localization.

SUVH3 promoter sequences

A 457 bp region extending from the 3′-end of the upstream gene (At1g73090) to the presumptive SUVH3 transcriptional start site (as discussed above; Fig. 1) was analysed for putative cis-acting regulatory elements using the PLACE database. Several putative transcription factor binding sites were identified in this region, most notably six DoF core motifs (5′-AAAG-3′), binding sites for members of the DoF transcription factor family that mediate light-induced, seed-specific and pathogen-inducible gene expression (Yanagisawa, 2004). The promoter region also contains two late pollen elements (5′-AGAAA-3′) shown to confer pollen-specific expression on the tomato gene LAT52.

Fig. 1. SUVH3 structure and nucleotide sequence of the 5′-upstream region, (A) Exons (E1–E3) are represented by rectangles and introns (i1, i2) by lines, 3′-g73 represents the 3′ end of the upstream gene (At1g73090). A bent arrow indicates the putative transcription start site and 5′ refers to the upstream promoter region. (B) The nucleotide sequence of the region between the 3′ end of the upstream gene (uppercase sequence) and the SUVH3 translational initiation (ATG) codon is shown. The transcriptional start site, based on the AY099620 cDNA is designated +1 and an asterisk marks the first nucleotide in the T04123 cDNA. DOF motifs, the LAT52 element, the as-1 elements, and a putative TATA box are shown in bold and italic letters. Within the transcribed region, exon and intron sequences are shown in upper case and lower case, respectively. The polypyrimidine (C/T) rich sequences in E1 and i1 are underlined.
(Twell et al., 1991) and a tandem repeat of the as-1 (activation sequence-1) motif 5′-TGACG-3′ (5′-TGACG-GCGGAGATGACG-3′; −139 to −122) originally identified in the CaMV 35S promoter (Lam et al., 1989). The as-1 motif mediates a variety of transcriptional patterns including expression in roots and leaves and in response to light, to salicylic acid and/or auxin and to abiotic and biotic stresses.

**SUVH3 is widely expressed during plant development**

In order to characterize the spatial (organ and tissue) and temporal expression patterns of SUVH3, a translational fusion was made between GUS and an approximately 2 kb region comprising 800 bp from the 3′ end of the adjacent upstream gene (At1g73090), the SUVH3 promoter region, the 5′-UTR, and the first five codons of the SUVH3 ORF. The resulting construct MAC-GUS was introduced into wild-type *A. thaliana* ecotype Columbia plants and the GUS expression patterns of eight independent transformed lines were determined by histochemical staining.

GUS activity was detected in most cell and tissue types (Fig. 3). The most intense GUS staining was observed in root tips with staining continuing into the root vascular tissue (Fig. 3A, B). The same expression pattern was observed in secondary roots. In leaves, GUS activity was detected in the mesophyll layer, in leaf veins, and in developing stomata and the epidermal cells surrounding them (Fig. 3C, D). GUS activity was also detected in stipules and in leaf primordia, but not in the shoot meristem.

In flowers, GUS was expressed in sepal, stamen filaments, mature pollen, and the gynoecium (Fig. 3E, F), and in the vascular tissue and the walls of green siliques (Fig. 3G). In the embryo, GUS activity was first detected at the globular stage, was maintained throughout embryo development but started to decline during embryo maturation and desiccation (Fig. 3H–K). During germination,
GUS activity was first detected in the emerging radicle (Fig. 3L) and as the seedlings developed, in the cotyledons (Fig. 3M, N). In root and leaves of seedlings the staining pattern was similar to that seen in mature plants (Fig. 3L-N).

**Intron 1 is a major determinant of high-level SUVH3–GUS expression**

In order to identify regulatory determinants of SUVH3 expression, several derivatives of the MAC–GUS fusion construct were generated (referred to as DEL constructs) that contained deletions of the region located upstream of the SUVH3 translational initiation codon (Fig. 4A). Because the resulting transgenic *Arabidopsis* lines contained one or more transgene insertions (data not shown), a logarithmic transformation was used to accommodate the large variation in GUS activity levels between transgenic lines containing a given construct. As shown in Fig. 4B, DEL2 plants showed levels of GUS activity similar to MAC–GUS plants (Tukey’s HSD test, *P*=0.531). The DEL2 construct carries a deletion of intron 2 suggesting
that it is not required for high-level GUS expression. DEL5 has a deletion that removes sequences derived from the 3' end of At1g73090, the gene located immediately upstream of the SUVH3 promoter (Fig. 4A). Although deletion of this region appears to reduce the level of GUS activity relative to that observed in MAC–GUS plants (Tukey’s HSD test, P=0.025), the levels in DEL5 plants are not significantly different from those observed in DEL2 plants (Tukey’s HSD, test P=0.464) (Fig. 4B).

In contrast to the above results, a dramatically reduced GUS activity was observed in DEL1 and DEL1-2 transgenic lines (Tukey’s HSD test, P < 0.001 and P < 0.001, respectively) (Fig. 4B). DEL1 plants in which the SUVH3 intron 1 was deleted showed a 190-fold reduction in GUS activity whereas DEL1-2 plants which carry a deletion of both introns 1 and 2, showed a 500-fold reduction in the levels of GUS activity. Transgenic plants containing the DEL4 construct showed only background GUS activity indicating complete loss of promoter activity. These results indicate that intron 1 is a key element regulating SUVH3 expression.

Surprisingly, during embryo development similar levels of GUS staining were observed in the MAC–GUS, DEL1, DEL2, DEL1-2, and DEL5 lines (Fig. 5D, I, N, S, X). Since each of these constructs contains the entire SUVH3 5'-flanking region, this clearly indicates that important regulatory elements also reside upstream of the 5'-UTR.

No GUS activity was detected in transgenic lines carrying the pDEL4 construct. This construct contains introns 1 and 2 but the 5' flanking region has been deleted and replaced with the –60 CaMV 35S minimal promoter which is itself unable to initiate transcription but can be transcriptionally activated by enhancer elements located nearby. The lack of detectable GUS activity in DEL4 lines suggests that the SUVH3 5'-UTR (and hence intron 1 and 2), does not contain enhancer elements capable of driving transcription independently of cis-acting elements located in the SUVH3 promoter region.

**Intron 1 is also a determinant of tissue- and developmental-specific expression patterns**

Histochemical staining of at least eight independent transgenic plants containing deletion constructs showed that GUS expression patterns conferred by pDEL2 (Fig. 5K–O) were identical to those conferred by the full-length pMAC–GUS construct (Fig. 5A–E) indicating that intron 2 does not regulate tissue-specific SUVH3 expression. Similarly, the pattern of GUS staining in DEL5 plants (Fig. 5U–Y) was identical to that observed in MAC–GUS lines although total GUS activity was lower in the DEL5 lines, especially in flowers.

In contrast to DEL2 and DEL5 plants, GUS staining patterns were identical in DEL1 (Fig. 5F–J) and DEL1-2 (Fig. 5P–T) plants, but differed from that observed in MAC–GUS lines. In roots of DEL1 and DEL1-2 lines, GUS was expressed in the columnella and the cells surrounding the quiescent centre, but not in the vascular tissue (Fig. 5F, P). GUS expression was also notably absent in flowers (Fig. 5H, R) and in leaf epidermal and mesophyll cells, was very
reduced in vascular tissue (Fig. 5G, Q) and was restricted to the root tips and the leaf primordia in seedlings (Fig. 5J, T). By contrast, in MAC–GUS, DEL2, and DEL5 lines, GUS activity was detected in all of these tissues and also in vascular tissue and in cotyledons. Since both DEL1 and DEL1-2 lack the first intron, this suggests that elements important for SUVH3 tissue-specific (root-, leaf-, and flower-specific) expression are located in this intron.

**Discussion**

The observed nuclear localization of the GFP-SUVH3 fusion protein in *Arabidopsis* and tobacco plants and cells was consistent with the presence in this protein of a bipartite NLS and an AT-hook domain that promotes binding to the minor groove of AT-rich tracts of DNA (Chuang and Kelly, 1999). GFP-SUVH3 accumulated exclusively in nuclei showing a diffuse distribution in the nucleoplasm...
and localization at discrete foci in interphase nuclei. Shen (2001) reported a similar pattern of subnuclear localization for NiSET1, a tobacco SUVH protein that belongs to the same orthologous group as SUVH3, and NiSET1, like SUVH3, remained associated with chromosomes throughout mitosis. More recently, immunodetection methods have localized the Arabidopsis SUVH1, SUVH2, and SUVH4 proteins in discrete foci which were shown to correspond to heterochromatic regions in interphase nuclei (Naumann et al., 2005; Fischer et al., 2006). This suggests that multiple subnuclear foci observed with the GFP-SUVH3 and NiSET1-GFP fusion proteins most likely represent heterochromatin or heterochromatin-like structures.

The evolution of multiple SUVH genes in plants, in contrast to animals, is thought to reflect fundamental differences in the regulation of developmental programmes at the level of chromatin modification. Development is considerably more plastic in plants than in animals and is profusely influenced by environmental factors. Moreover, unlike in animals, organ development in plants is not restricted to the embryonic stage, but occurs throughout the life-cycle. Functional differences between plant SUVH genes might be based on the specific activities of the encoded proteins and/or differential regulation of SUVH gene expression during the plant life-cycle. It has been shown, for example, that SUVH4 and SUVH2 (Jackson et al., 2004; Naumann et al., 2005), but not SUVH1, SUVH5, or SUVH6 (Naumann et al., 2005; Ebbs and Bender, 2006) play major roles in the organization of heterochromatin in Arabidopsis. Thus far, expression analyses based on data derived from RNA gel blots and RT-PCR have concluded that SUVH genes are expressed constitutively in all organs and tissues. However, the analysis of SUVH3-GUS gene expression patterns clearly shows that although GUS activity was observed in most plant tissues, the levels varied greatly between tissues.

Deletion analysis of the region upstream of the SUVH3 translational start codon identified intron 1 as a key regulator of SUVH3-GUS expression except in embryos where the levels of GUS activity appear to be independent of its presence/absence. The enhancement of gene expression by introns may be related to transcription initiation when specific transcriptional enhancer elements are present in the intron, as has been shown for the second intron of the Arabidopsis gene AGAMOUS (Deyholos and Sieburth, 2000). Alternatively, introns can enhance gene expression by increasing mRNA accumulation without significantly affecting the rate of transcription initiation (Rose, 2002; Rose and Last, 1997). Intron 1 of SUVH3 did not direct significant levels of GUS expression when fused with the -60 CaMV 35S minimal promoter. This observation suggests that intron 1 does not contain a cryptic promoter or enhancer elements capable of activating transcription from a minimal promoter. It does not, however, rule out the possibility that intron 1 may contain enhancer-like elements capable of interacting in a combinatorial manner with enhancers located in the SUVH3 5′-upstream region. Indeed the large decrease in GUS activity in DEL1 plants (which retain intron 2), but not in DEL2 plants (which retain intron 1), relative to MAC–GUS plants suggests the presence of enhancer-like elements in intron 1.

Constitutive, high-level expression of the Arabidopsis replacement histone H3 genes has also been shown to depend on an intron located in the 5′-UTR (Chaubet-Gigot et al., 2001). Neither the intron nor the 5′-UTR contains cryptic promoter sequences. In alfalfa, the 5′-UTRs of the three constitutively and highly expressed replacement H3 genes lack introns, but in common with the intron-containing 5′-UTRs of the Arabidopsis histone H3 genes, contain a high density of GAGA-like elements (Robertson et al., 1996; Chaubet-Gigot et al., 2001). Intragenic GAGA elements comprising repeats of the dinucleotide sequence (GA)n or (CT)n occur in approximately 6% of Arabidopsis genes (Meister et al., 2004). In Drosophila, GAGA elements have been shown to repress the expression of numerous developmental genes by enhancing the stability of nucleosomes. Repression is relieved by a protein called the GAGA factor, encoded by the trithorax-like gene, which binds to the GAGA element and facilitates local disruption of nucleosomes (Tsukiyama et al., 1994). In plants, GAGA elements are bound by members of the BASIC PENTACYSTEINE (BPC) protein family that are not related to the Drosophila GAGA factor (Sangwan and O’Brian, 2002; Meister et al., 2004). Although the regulatory consequences of BPC binding to GAGA elements is at present unclear, the evidence supports a positive rather than a negative role in the regulation of gene expression. Thus, the GAGA-like elements located in exon 1 and intron 1 of the SUVH3 5′-UTR, immediately adjacent to the promoter region, may facilitate constitutive SUVH3 expression by excluding nucleosomes and other repressive factors (Chaubet-Gigot et al., 2001).

Transgenic plants expressing the DEL1-2 derivative showed a dramatically reduced GUS activity except in the embryo. This expression pattern, conferred by the SUVH3 promoter region alone, is strikingly similar to that observed in tobacco plants expressing GUS under the control of subdomain AI (–90 to +8) of the CaMV 35S promoter (Benfey et al., 1990). Both promoters confer expression in root tips and weak expression in the meristematic region of the seedlings and in the vascular tissue of stem and leaves. Neither promoter drives expression in flowers and only very weak expression in the floral stalk. The cis-element as-1 (activation sequence 1), a tandem repeat of TGACG, which binds the nuclear factors ASF-1 (activation sequence factor 1) and the transcription factor TGA1a (Katagiri et al., 1989; Lam et al., 1989) has been shown to be responsible for this expression pattern in the –90 CaMV 35S promoter. Interestingly a similar motif (TGACGGCGGAGATGACG) is found in the SUVH3
promoter, which may account for the expression pattern conferred by the $SUHV3$ promoter in the absence of the downstream 5'-UTR introns.

Since no change in expression was observed in embryos for any of the $SUHV3–GUS$ constructs lacking intron 1 and/or intron 2, critical cis-acting elements responsible for embryo-specific expression are probably located in the 5' upstream region. However, no known cis-element related to embryo expression was found. This suggests that unidentified cis elements or a so far unidentified combination of motifs may confer the observed expression of $SUHV3$ in embryos.

**Acknowledgements**

NTL was supported by Enterprise Ireland Grant SC/2002/486. JAC-M was supported by EU contract QLG 2-1999-00876 (REGIA).

**References**


